Ratings as validation of sociometric status determined by nominations in longitudinal research

Maassen, G.H.; Goossens, F.A.; Bokhorst, J.

published in
Social Behavior and Personality
1998

DOI (link to publisher)
10.2224/sbp.1998.26.3.259

document version
Publisher's PDF, also known as Version of record

SOCIAL BEHAVIOR AND PERSONALITY, 1998, 26(3),259-274

RATINGS AS VALIDATION OF SOCIOMETRIC STATUS DETERMINED BY NOMINATIONS IN LONGITUDINAL RESEARCH

GERARD H. MAASSEN
Utrecht University, The Netherlands
FRITS A. GOOSSENS AND JACOBUS BOKHORST
Free University of Amsterdam, The Netherlands

The current methods for two-dimensional sociometric status classification into five categories are based on nominations. Recently, Maassen et al. (1996) showed that such a classification can also be achieved by means of rating-scales. Ratings have been known for some time to be more reliable than nominations, an important consideration in longitudinal research. We report an investigation among school pupils, who were assessed by their classmates by means of nominations at three moments, with an interval of one year between the measurements. The sample consisted of 157 children with a mean age of 7 years and 7 months at the third moment. We demonstrate how status passages assessed by means of nominations can be validated by an extra assessment with ratings. Particular attention is paid to passages from and into the category of rejected children.

The determination of sociometric status within a peer group is an important element of many developmental studies. One generally assumes that peer status is a proxy variable for social competence and a useful predictor for adjustment in later life. The determination of sociometric status is usually aimed at identifying those youths who have problems relating to their peers (Asher & Coie, 1990; Cillessen, van IJzendoorn, Van Lieshout & Hartup, 1992; DeRosier, Kupersmidt

Gerard H. Maassen, Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University; Frits A. Goossens, Department of Education, Faculty of Psychology and Education, Free University of Amsterdam; and Jacobus Bokhorst, Department of Education, Faculty of Psychology and Education, Free University of Amsterdam
Please address correspondence and reprint requests to: dr. Gerard H. Maassen, Department of Methodology and Statistics, F.S.W. Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands, fax: +31 30 2546530. (Email: g.maassen@fsw.ruu.nl)

& Patterson, 1994). Further research amongst such children is important because they run a higher risk of maladjustment later on in life.

Until the early eighties, sociometric status was assessed in a one-dimensional way, employing either nominations or ratings. The nominations-based procedure consists of asking every child from a group (usually a classroom group) to nominate some peers with whom they would like to undertake a certain activity (e.g. playing), whilst at the same time the children are also asked to nominate peers with whom they would least like to undertake that same activity. The ratings-based method consists of asking every child from a group to indicate on a rating-scale how much they like or dislike every other child from the group. Both procedures aimed at a one-dimensional classification of social acceptance of individuals within their peer group.

Already at an early stage the idea of a more refined determination of sociometric status was expressed (Bronfenbrenner, 1944; Dunnington, 1957). Researchers needed (Gronlund & Anderson, 1957) not only to be able to identify the *rejected* pupils, but also to identify those who by their very inconspicuousness have few contacts with their peers (*neglected* pupils). Peery (1979) was the first to suggest a practical two-dimensional procedure. On the basis of his proposals, Coie and Dodge (Coie, Dodge & Coppotelli, 1982; Coie & Dodge, 1983) and Newcomb and Bukowski (1983) developed new procedures using nominations which lead to a classification into five different sociometric status groups: *popular*, *rejected*, *neglected*, *controversial* and *average*. This distinction is considered to be two-dimensional; popular, average and rejected correspond to the first dimension (referred to as *social preference*), while neglected, average and controversial correspond to the second dimension, usually described as *social impact*. The procedures of Coie et al. and of Newcomb and Bukowski differ only in the statistical rules for allocating subjects to a certain status group. Both are at present considered appropriate methods for determining sociometric status and have been applied in numerous studies (see Newcomb, Bukowski & Pattee, 1993). The Coie et al. procedure, however, appears to be used more often, probably because its statistics are somewhat easier to adopt in actual research.

Rating-scales continued to be applied in the determination of sociometric status (French, 1988, 1990), but their use has strongly declined since the beginning of the eighties. Rating-scales appeared not to lend themselves readily to a classification into the five different categories which had since become popular. In addition, using nominations had the distinct advantage of giving respondents an easier task: mentioning only some liked and disliked peers rather than assigning scores to *all* other pupils of the group. This makes a nominations-based system more appropriate for large-sized groups or for groups where not all members are known to each other.

On the other hand, ratings have their advantages, too. We have recently pointed out that there is a good conceptual argument for the use of rating-scales rather than nominations (Maassen, Van der Linden & Akkermans, 1997). When using nominations, in principle every respondent has to mention a prescribed number of group members (usually three) as 'least liked' or 'most liked'. (The critical value in the Newcomb and Bukowski procedure is based on such fixed numbers.) However, it is possible that the respondent likes all the others to some extent, in which case the nomination procedure forces the respondent to be negative about some other group members despite the respondent's actual positive feelings. In practice, this drawback is mitigated since in most research it is tolerated for respondents to mention fewer than three nominees.

Yet another disadvantage of nominations is that only limited information can be collected and processed. The sympathy or antipathy one feels for the nominated persons is likely to differ, but these differences cannot be made explicit with nominations. The same can be said with regard to the non-nominated group members. Rating-scales enable the researcher to collect information which is more refined and better reflects the respondent's actual feelings about his or her fellow group members. This makes the ratings method particularly apt for moderate-sized groups where members know each other fairly well (classroom groups for instance). Moreover, the greater variability of the rating scores makes them more appropriate for follow-up analyses of a more advanced type (such as multiple regression or multidimensional scaling).

Following these arguments, Maassen, Akkermans and Van der Linden (1996) developed a method which allows the transformation of rating-scale data into a two-dimensional sociometric status assessment, based on the five categories mentioned above. The method has been concretized in a corresponding computer program (SSRAT)[1]. This development makes the nominations and the ratings methods more comparable, particularly with respect to the issue that is central to this article and to which we will now turn: longitudinal stability.

Before the introduction of the nomination procedures mentioned above, when sociometric status assessment was still one-dimensional, it was demonstrated that the test-retest reliability of peer ratings was superior to that of nominations (Asher & Hymel, 1981). High reliability, of course, is expected to enhance longitudinal stability (when actual changes in peer status are absent). On this point, however, the literature is not consistent. Putallaz and Gottman (1981) contended that ratings provide better stability, but Terry and Coie (1991) found no difference

[1] The program and manual (Maassen & Landsheer, 1996) are available from the first author.

between the stability of nominations-based and one-dimensional ratings-based measures. Terry and Coie (1991) conclude: "that groups formed on the basis of a one-dimensional variable, such as ratings or social preference, will have less desirable psychometric properties than those formed by a two-dimensional system." Since SSRAT is available, which allows a two-dimensional processing of ratings (i.e. a separate treatment of the extreme categories of neglected and controversial persons), their argument against the ratings method is no longer relevant.

In the present investigation the stability of sociometric status determination has been examined anew. A full-scale comparison between the nomination procedures and the two-dimensional ratings method with regard to their stability, however, is not yet at issue, since it will take several years before the required data are collected. The aim of this study is to show that the ratings procedure can be a useful tool for the validation of certain status transitions which may occur in longitudinal research. For instance, a researcher may have assessed the sociometric status of a number of subjects at two different points in time and may wish to know which subjects have moved to another sociometric status group. It is likely that he is specifically interested in those individuals who have moved into or out of the rejected group. If the researcher is used to assess sociometric status by means of nominations, he may expand his most recent 'wave' with rating-scale data and thus be in the position to validate the nominations-based sociometric status transitions. This paper addresses this situation and its main research question can be formulated as follows: How can rating-scale scores collected at T2 be used to validate sociometric status transitions assessed by nominations at T1 and T2?

METHOD

The Sample

The study commenced in May 1994 (T0) with 186 pupils, nearly all the children from two protestant Christian schools in Badhoevedorp and Hoofddorp, two suburbs close to Amsterdam. The level of non-response was less than 5%. Pupils were divided over 6 different classrooms. Every child was asked to express his or her sympathy or antipathy for the other members of the class by way of nominations. This was repeated in May 1995 (T1) and May 1996 (T2). On the last occasion, rating-scales were also used. At each point in time, the composition of the classrooms differed from the time before: some children had moved, others were new, and the distribution over the classrooms had changed.

The nomination data at the first point in time (T0) turned out to be rather unreliable, presumably because of the age of the children at that time (5 years and 4 months; education is available from age 4 and mandatory from age 5 in the

Netherlands). The young children often nominated the same persons as most liked and most disliked. For this reason, we limited ourselves to the data collected at T1 and at T2. For this paper, we use the data from those pupils who were in the sample at all three times of measurement and from whom nominations were collected at T1 and both nominations and ratings at T2. This amounts to a total of 157 pupils (77 girls; 49%). At T2 the ages ranged from 6 years, 7 months to 8 years, 10 months. The average age was 7 years and 7 months (N=152; the age of five pupils was unknown). The classroom sizes (in parentheses the number of pupils taking part in the study) at T2 were: 28 (23), 27 (26), 32 (27), 30 (29), 28 (26) and 31 (26).

CLASSIFICATION INTO STATUS CATEGORIES USING RATINGS AND NOMINATIONS.

To enhance understanding of the following we first describe the essentials of the nominations and the ratings procedures. At present, in most of the investigations where sociometric status of persons is determined by means of nominations, the classification procedure of Coie et al. is used. We therefore confine ourselves to this variant, which will be referred to as *CDCnom*.

In our research, the sample consists of school pupils, and the classroom groups to which they belong are the social groups to be distinguished. Within the framework of CDCnom each pupil is asked two questions: (1) "Which classmates do you like most?", and (2) "Which classmates do you like least?". As usual in this type of research the respondents had to nominate three persons in both categories. Let variable *LM* (Liked Most) be the total number of positive nominations received by a person and *LL* (Liked Least) the total number of negative nominations. *LM* and *LL* are standardized within each classroom group. Subsequently, scores for social preference and social impact are determined as the difference and the sum, respectively, of these standard scores (thus $SP = z_{LM} - z_{LL}$ and $SI = z_{LM} + z_{LL}$) and standardized as well. Classification into sociometric status groups according to CDCnom is then as follows:

1.	popular:	$z_{SP} > 1$,	$z_{LM} > 0$ and $z_{LL} < 0$;
2.	rejected:	$z_{SP} < -1$,	$z_{LM} < 0$ and $z_{LL} > 0$;
3.	neglected:	$z_{SI} < -1$,	$z_{LM} < 0$ and $z_{LL} < 0$;
4.	controversial:	$z_{SI} > 1$,	$z_{LM} > 0$ and $z_{LL} > 0$;
5.	average:	the remaining group members.	

Within the framework of the ratings method, *SSrat* (from now on), each pupil is asked to rate each classmate on a scale ranging from *dislike very much* to *like very much*. For the data collection (2*R*+1)-point rating-scales are used, i.e. scales with an odd number of scale points; the scale's midpoint must represent a neutral judgment. (In our research, where a 7-point scale was used, *R*=3). Data are

arranged in a matrix **P**, with rows belonging to assessors and columns to persons assessed. (Let P_{ik} denote the rating given by assessor i to group member k.)

In SSrat the criteria for classification into sociometric status groups are applied to statistics derived from **P**. First, R+1 is subtracted from all scores in **P**; the resulting matrix **P*** contains the elements $-R,\ldots, 0,\ldots, R$. This matrix is then transformed into a matrix **I** of *impact-scores* I_{ik} by taking the absolute value for all the elements in **P***. A matrix **S** of non-negative scores S_{ik} (*sympathy-ratings*) is created by substituting 0 for all the negative elements of **P***. Analogously, a matrix **A** of non-negative *antipathy-ratings* A_{ik} is derived by substituting 0 for all the positive elements of **P*** and by taking the absolute value for all the negative elements. Then, for the judgment of an individual assessor i with respect to person k:

$P^*_{ik} = S_{ik} - A_{ik}$ and $I_{ik} = S_{ik} + A_{ik}$,

and consequently for the judgment at group level with respect to person k:

$$P^*_{+k} = \Sigma\ P^*_{ik} = \Sigma\ S_{ik} - \Sigma\ A_{ik} = S_{+k} - A_{+k}$$

$$I_{+k} = S_{+k} + A_{+k}$$

(Note the analogy with the definitions of social preference and social impact by Coie et al.)

In SSrat, too, classification is predicated on the levels of the observed values of P^*_{+k} and I_{+k}. For each person assessed, a probability distribution for the possible values of P_{+k} (which is identical to the probability distribution of P^*_{+k}, taking the correspondence between the two sets of scores into account) is determined theoretically and estimated in practice. The determination of this distribution is based on the null hypothesis of conditional random score attribution by the assessors, a principle that has been applied before by Ten Brink (1985). This means that for each assessor a different multinomial probability distribution of the attributed ratings is estimated. Then, the distribution of P_{+k} becomes the sum of these multinomially distributed random variables. In a similar way, the probability distributions of I_{+k} (social impact), S_{+k} (sympathy ratings total) and A_{+k} (antipathy ratings total) of each assessed person can be determined. (For a mathematically more detailed description of the calculation of these probability distributions, the reader is referred to Maassen et al., 1996). The procedure implies that the attributed ratings are weighted differently: A high rating received from a person who attributes almost only high scores is weighted less heavily than a high rating from a person who scarcely attributes high scores. As a consequence, it is possible that two persons who received the same ratings total are classified differently.

In order to achieve a classification into sociometric status groups, the actual values of these statistics need to be checked against criteria. The criteria of SSrat basically are a translation of the criteria of Coie et al. into probability terms:

1. popular: P_{+k} significantly high, $S_{+k} > E(S_{+k})$ and $A_{+k} < E(A_{+k})$[2];
2. rejected: P_{+k} significantly low, $S_{+k} < E(S_{+k})$ and $A_{+k} > E(A_{+k})$;
3. neglected: I_{+k} significantly low, $A_{+k} < E(A_{+k})$ and $S_{+k} < E(S_{+k})$;
4. controversial: I_{+k} significantly high, $A_{+k} > E(A_{+k})$ and $S_{+k} > E(S_{+k})$;
5. average: the remaining group members.

PROCEDURE

At T1 (nominations) all children were tested individually by a female experimenter in a quiet room inside the school building. A photograph was shown of all children making up the class, and the children were first asked to identify themselves and all others in the photograph before the experimenter asked them to nominate three children with whom they would like to play. She then checked by asking the children to go over the photograph just "to see if there were not any children the subject had overlooked and would rather substitute for those nominated." Next, the procedure was repeated for the negative nominations.

At T2 (nominations and ratings) the procedure for the nominations was altered somewhat. The children's teachers served as the experimenters. The teachers supplied all children with a list of names of all the pupils, assured them that no information would be revealed to any of the other children, and then asked them to nominate three positive peers ("someone you like to play with"). When they had all done so, they were asked to check the list of names again and decide whether they wanted to change any name. Subsequently, the procedure was repeated for the negative nominations.

For the ratings procedure, the children were all tested individually by a female experimenter in a quiet room inside the school building. They were presented with a randomized list of all the names of the other classmates (randomization was performed for each individual child) and asked to indicate on a 7-point scale (1 = *dislike strongly*, 4 = *neutral*, 7 = *like very much*) how they felt about each of their classmates. The experimenter first presented the child with an example, checked whether the child had understood the instructions and then went through the list with the child.

[2] Here E denotes the expected value operator on a probability distribution.

RESULTS

First the classification results of the CDCnom procedure at T1 and T2 (CDCnom1 and CDCnom2) as well as the ratings procedure SSrat at only T2 are presented and discussed. The size of the categories yielded by a classification method depends on the cut-off points chosen within the framework of the procedure used by the researcher. Terry and Coie (1991) state: "The question of appropriate size depends on the purpose group selection is to serve." For the moment, possible purposes of group selection are not relevant to our argument. With reference to the CDCnom procedure, the cutting score is usually set at one standard deviation from the mean and this was the choice made when Table 1 was prepared. In order to determine an appropriate cutting criterion within the SSrat framework, the characteristics of a series of possible classifications are examined first. SSrat (like the Newcomb and Bukowski procedure) is a probability method, i.e. classification into one of the five sociometric status groups depends on a probability criterion set by the researcher. The table presents the distributions over the sociometric status groups according to a number of values of parameter alpha (significance level). This is done only by way of illustrating our argument; usually the researcher will select one or two levels and stick to it[3].

TABLE 1
CLASSIFICATION INTO FIVE SOCIOMETRIC STATUS GROUPS (IN PERCENTAGES) ACCORDING TO SSRAT WITH VARIOUS LEVELS OF SIGNIFICANCE AT T2 AND ACCORDING TO THE METHOD OF COIE ET AL. AT T1 (CDCNOM1) AND T2 (CDCNOM2) (N = 157)

	popular	average	neglected	controvers.	rejected
Method					
SSRAT.050	29	46	4	0	21
SSRAT.025	23	57	2	0	18
SSRAT.010	15	67	1	0	17
SSRAT.005	11	73	1	0	16
SSRAT.004	10	73	1	0	16
SSRAT.003	10	75	1	0	14
SSRAT.002	6	81	1	0	12
SSRAT.001	4	85	0	0	11
CDCnom1	13	57	12	6	12
CDCnom2	13	60	9	5	13

[3] Maassen et al. (1996) recommend using .025. On the basis of recent studies Maassen, Van der Linden, Goossens and Bokhorst (forthcoming) advise: "When the categories 'neglected' and 'rejected' are important to the research, an α value not lower than .05 should be used. When the categories 'popular' or 'rejected' are central to the research, an α of .01 performs well. A combination of both variants may be used if all five sociometric status categories are important."

The table shows clear differences between the nomination and the rating results, which also showed up in a previous study (Maassen et al., 1997). SSRAT did not assign a single pupil to the controversial status when using an alpha of .05 or less, while under these conditions the percentage of neglected children was considerably lower with SSrat than with CDCnom. We would like to comment on this phenomenon.

SSrat is certainly able to detect outspoken, neglected or controversial individuals (see Maassen et al., 1996, Tables 1 and 2). However, it was also found in previous research that SSrat yields considerably lower numbers of neglected and controversial persons than the nominations procedures (see Maassen et al., 1996, Table 3). SSrat is more discriminating in denominating persons as neglected or controversial, which is readily understood. Being neglected according to SSrat, for instance, means that many fellow group members expressed no degree of like or dislike of the person in question, while in the framework of the nominations procedure such a person was not mentioned by many group mates as either liked least or liked most, which naturally occurs more often. Technically, this implies that sometimes a significantly low or high impact score according to SSRAT is rarely found, whereas the fixed cutting score in CDCnom results in almost fixed-sized categories of neglected or controversial persons. (If the research goals make higher numbers of neglected or controversial persons desirable, the researcher who applies SSRAT may of course increase the alpha value along the impact dimension.) The very low numbers of neglected and controversial children in our study, however, suggest that a one-dimensional model of sociometric status might be adequate in this case.

Other remarkable differences concern the popular and the rejected children. At alpha = .05, SSRAT assigns much larger percentages of the pupils to the popular or rejected status groups than does CDCnom. Choosing sufficiently low alphas reduces their percentages to below that of CDCnom.

Table 2 shows the composition of various groups of popular and rejected children using different alpha levels. The comparison between SSrat and CDCnom is based on the data at T2 when both nominations and ratings were collected. In this comparison the average rating received by a pupil from all his classroom peers (variable *arr*) plays a central role. For every alpha level chosen, the mean, the standard deviation, the highest and the lowest score on this variable have been given. In addition, an out-of-range criterion has been applied, which counts the number of anomalies of a distribution. As anomalies we consider: (1) a pupil assigned to the popular status group with an *arr* lower than the neutral scale point (i.e. 4.0), and (2) a pupil assigned to the rejected status with an *arr* higher than the neutral scale point. With variable *arr* and the out-of-range criterion we can get an indication of the quality of the various distributions.

With regard to the table, we would like to point out that SSRAT with alpha equal to .05 produces a considerably larger 'popular' group than the CDCnom procedure. Nevertheless, its *arr*-characteristics are more favourable: a higher mean and a higher minimum score. Lowering the alpha level makes the distribution more selective: the size decreases while the *arr*-characteristics become even more favourable. With an alpha level of .007, the size of the 'popular' group according to SSrat is almost the same as that on the basis of the CDCnom procedure.

TABLE 2

DATA OF VARIABLE ARR (AVERAGE RECEIVED RATING) FOR POPULAR AND REJECTED PUPILS ACCORDING TO SSRAT WITH VARIOUS SIGNIFICANCE LEVELS AND ACCORDING TO THE METHOD OF COIE ET AL. AT T2

method	M	SD	min	max	N	perc	out of range[a]
populars							
SSRAT.050	5.25	0.28	4.73	5.81	45	29	0
SSRAT.025	5.29	0.27	4.81	5.81	36	23	0
SSRAT.010	5.38	0.27	4.96	5.81	24	15	0
SSRAT.005	5.40	0.28	5.04	5.81	17	11	0
SSRAT.004	5.42	0.28	5.04	5.81	16	10	0
SSRAT.003	5.42	0.28	5.04	5.81	16	10	0
SSRAT.002	5.52	0.28	5.12	5.81	10	6	0
SSRAT.001	5.58	0.25	5.19	5.81	6	4	0
CDCnom2	5.48	0.30	4.43	5.81	20	13	0
rejected							
SSRAT.050	3.35	0.47	2.12	4.35	33	21	3
SSRAT.025	3.28	0.45	2.12	4.23	29	18	2
SSRAT.010	3.24	0.42	2.12	4.08	27	17	1
SSRAT.005	3.22	0.42	2.12	4.08	25	16	1
SSRAT.004	3.22	0.43	2.12	4.08	25	16	1
SSRAT.003	3.15	0.39	2.12	3.76	22	14	0
SSRAT.002	3.10	0.40	2.12	3.72	19	12	0
SSRAT.001	3.07	0.38	2.12	3.58	18	11	0
CDCnom2	3.36	0.53	2.12	4.39	21	13	0

[a] For popular *arr* > 4.0; for rejected *arr* < 4.0

We would argue similarly for the rejected group. Using an alpha level of .05, SSrat assigns considerably more children to the rejected group than does the CDCnom procedure. The *arr*-characteristics of both distributions are similar, but the distribution according to SSrat shows more anomalies. The characteristics of the distribution according to SSrat can be improved by choosing a lower alpha level. With an alpha of .003, the size of the rejected group according to SSrat is about the same as that of the CDCnom procedure. Then the *arr* has a lower mean, a lower maximum score and no longer contains any anomalies.

The main research question, formulated earlier, concerns the extent to which ratings can be used to validate status transitions which were determined by way of nominations. The ratings procedure is particularly useful as a validation instrument if the longitudinal stability of SSrat is superior to that of a nominations

procedure as CDCnom. Stability of a method depends on the sociometric status category involved and the number of persons classified into this category. Because the rejected group is very often the central object of research, we shall pay special attention to the distribution according to SSRAT with alpha = .003 (SSRAT.003, from now on). By choosing alpha equal to .003, we prevent SSrat becoming more stable (with respect to the rejected group) solely as a consequence of larger group size.

In order to gain insight into the longitudinal stability of the nominations and the ratings procedure, we will now compare the various distributions of the sociometric status groups at the two moments of measurement. Three comparisons can be made: CDCnom1 with CDCnom2, CDCnom1 with SSRAT.003 and CDCnom2 with SSRAT.003. Cross-tabulations of the first two comparisons are presented in Table 3.

TABLE 3

CLASSIFICATION INTO SOCIOMETRIC STATUS GROUPS (ABSOLUTE FREQUENCIES) AT T1 AND T2 ACCORDING TO TWO WAYS OF PERFORMING REPEATED MEASUREMENTS: NOMINATIONS-NOMINATIONS AND (BETWEEN PARENTHESES) NOMINATIONS-RATINGS

	CDCnom1					
	pop	ave	neg	con	rej	total
CDCnom2 (SSRAT.003)						
popular	4 (6)	11 (9)	2 (1)	3 (0)		20 (16)
average	15 (14)	55 (73)	12 (17)	6 (8)	6 (6)	94 (118)
neglected	1 (0)	8 (1)	4 (0)		1 (0)	14 (1)
controversial		5 (0)		1 (0)	2 (0)	8 (0)
rejected		10 (6)	1 (1)	0 (2)	10 (13)	21 (22)
total	20	89	19	10	19	157

In our sample, the distributions according to CDCnom2 and SSrat showed a considerably higher agreement than those of CDCnom1 and CDCnom2. In the first comparison, 103 pupils (66%) received the same sociometric status classification (Cohen's kappa = .36), whereas in the second comparison only 74 pupils (47%) received the same classification (Cohen's kappa = .14). This is not surprising, as CDCnom1 took place at a different point in time than both CDCnom2 and SSrat, which took place at the same moment. Another result of these comparisons is more remarkable: the stability of the sociometric status classifications by way of nominations (CDCnom1 and CDCnom2) was also lower than the stability by way of nominations at T1 and ratings at T2. In the first comparison only

74 pupils (47%) remained in the same status group, whereas in the second comparison 92 pupils (59%) remained in the same status group (Cohen's kappa = .24).

Usually, researchers are interested in transitions to and from the 'rejected' group. Table 4 presents these transitions according to both procedures (CDCnom1-CDCnom2 and CDCnom1-SSRAT.003): improvement of sociometric status (transition *from* the rejected group), deterioration of sociometric status (transition *into* the rejected group), and also unchanged status (stationary in the rejected group), etcetera.

Table 4
STATUS PASSAGES INTO AND FROM THE REJECTED CATEGORY FOR TWO COMBINATIONS OF REPEATED MEASUREMENT (NOMINATIONS-NOMINATIONS AND NOMINATIONS-RATINGS)

	nominations-nominations				
	into rej.	from rej.	station. rej.	outside rej.	total
nominations-SSRAT.003					
into rej.	6			3	9
from rej.		5	1		6
station rej.		4	9		13
outside rej.	5			124	129
total	11	9	10	127	157

There were 33 transitions which, in one way or another (at one or both points in time with either one or both procedures), involved the rejected group. In more than half of these cases (20 pupils) both procedures led to a similar type of status transition. That is, almost all the children who remained in the rejected group according to the CDCnom1-CDCnom2 procedures also remained rejected when different classification procedures at the two times of measurement were used (CDCnom1-SSRAT.003). However, in the remaining cases caution is needed.

In about half of the cases where, according to the CDCnom1-CDCnom2, children improved their sociometric status by moving out of the rejected group, this was not supported by the rating procedure. The same result (no support from the rating procedure) could be observed for those children who according to the dual nomination procedure had moved into the rejected group. Finally, using the rating procedure would suggest that certain children had declined in sociometric status, but this would not have been detected by repeated measurements involving nominations.

SUMMARY AND DISCUSSION

A central issue of this article is the stability of social status determination, i.e. the capability of current procedures to produce the same results at different moments when no actual change in social status has occurred. In particular, we focus attention on the difference in stability between the nominations-based and the ratings-based methods. When sociometric status determination was still one-dimensional (i.e. until the early 1980s) several authors contended that the test-retest reliability of peer ratings was superior to that of nominations (Asher & Hymel, 1981) and that ratings provide better stability (Putallaz & Gottman, 1981). Since the beginning of the 1980s, when nominations became employed for two-dimensional classifications of sociometric status, a sound comparison between the two methods with regard to their stability was no longer possible. Terry and Coie's (1991) conclusion that groups formed on the basis of a one-dimensional rating-scale will have less desirable psychometric properties than those formed by the two-dimensional nominations procedures, placed the ratings method in a poorer position.

The availability of the ratings-based method SSRAT, which enables the researcher to employ also ratings for two-dimensional classifications, revives the question of whether there is a difference in stability. In order to answer this question, a large-scale longitudinal study is required in which the results of sociometric classifications based on ratings are compared with those based on nominations. The efforts of the present authors in this field will continue for several years before more decisive results can be presented.

The present study, however, which is based on a more restricted design, shows a remarkable support for our contention that, *ceteris paribus*, the ratings method provides superior stability. This study has involved a measurement by way of nominations at times T1 and T2, and a measurement by way of ratings only at time T2. Classifications on the basis of repeated measurements by way of nominations showed a lower stability than the two classifications using nominations at time T1 and ratings at time T2. This result is best seen as yet another indication of the higher reliability of the rating-scale scores.

Now that SSRAT enables the researcher to obtain two-dimensional classifications based on ratings, it is of interest to confront this finding with the conclusion of Terry and Coie cited earlier. The classification according to SSRAT, which yields hardly any neglected or controversial children, suggests that the data of the present study are probably adequately represented by a one-dimensional model. (With reference to the classifications according to the nominations procedure, we note that the attributions to these categories are highly unstable: only four of the nineteen individuals originally denominated as neglected and one of the ten individuals originally denominated as controversial remain in the same category at

time T2.) This would seem to indicate that it is not the dimensionality of the ratings procedure that matters, but the collection and processing of more refined information.

If the findings suggest more favorable psychometric properties for the ratings-based method, ratings collected at the most recent time of measurement can play a role as a validation instrument. It should be emphasized that the method described in this study is to be regarded as just an example in several respects. Several nominations-based procedures are available, as well as an infinite number of options for the cut-off criterion adopted within the method to discriminate between the extreme categories and the average group. We assume that the researcher who wishes to apply the proposed procedure, will have made his own choice with respect to the nominations method (including its cutting criterion) that he has used in the past. For the example in the present study, the most commonly used nominations method was chosen (that of Coie et al. with a cut-off point at one SD from the mean).

We have focused on status transitions involving the group of rejected children. In many cases, researchers can increase their confidence in the results because the outcomes of the nominations method and SSRAT are convergent, especially in the case of children who remain in the rejected status group. In many of the remaining cases (where one might wrongly conclude that some child has either deteriorated or improved in status), our results show that researchers should be cautious before leaping to such conclusions. How researchers should deal with this is a rather subjective matter. We have strong confidence in the reliability of transitions that have been established by way of both nominations and ratings, but less confidence in transitions that have been established on the basis of nominations used at two points in time but have not been supported by the trajectory nominations followed by ratings.

In our validation we used SSRAT with a relatively low value of alpha (.003). One may wonder to what extent the validation results are dependent on the choice of alpha. Choosing a higher value for this parameter places more pupils in the rejected group. In other words, the CDCnom1-SSrat trajectory will then show fewer transitions from the rejected group, more transitions into the rejected group and more children who remain in that group. The validation results are not independent of alpha. Since the central question of this paper assumes (1) the longitudinal use of nominations as a basic element of a study and the collection of ratings for validation purposes only, and (2) that transitions into and from the rejected group are the focus of attention, it would seem reasonable to choose an alpha value which leads to a similar number of rejected pupils under both procedures.

One may well ask whether SSRAT is also suitable for the validation of other types of status transitions. In quite a few studies the group of popular children

figure as a reference group. SSRAT can also be used as a validation measure for the composition of this group. To be able to delimitate the 'truly' popular children, it is important that these children should be classified as such at more than one time. The following procedure can then be used. Assess how many children are classified as 'popular' by both CDCnom1 and CDCnom2 and validate this group by way of SSRAT choosing an alpha level which leads to a similar sized group of popular children. Those children who are included in the popular status group according to both methods are the truly popular ones.

It should be apparent from the above (see Results) that SSRAT cannot be used for the validation of transitions into or from the categories of 'neglected' and 'controversial', since these will occur far less often when using ratings.

REFERENCES

Asher, S.R., & Coie, J.D. (Eds.). (1990). *Peer rejection in childhood*. New York: Cambridge University Press.

Asher, S.R., & Dodge, K. A. (1986). Identifying children who are rejected by their peers. *Developmental Psychology*, **22**, 444-449.

Asher, S.R., & Hymel, S. (1981). Children's social competence in peer relations: Sociometric and behavorial assessment. In J.D. Wine & M.D. Smye (eds.), *Social competence*, 125-157. New York: Guilford.

Bronfenbrenner, U. (1944). A constant frame of reference for sociometric research. Part II: Experiment and inference. *Sociometry*, **7**, 40-75.

Cillessen, A.H., & Ten Brink, P.W.M. (1991). Vaststelling van relaties met leeftijdgenoten. (Determining relationships among peers) *Pedagogische Studiën*, **68**, 1-14.

Cillessen, A.H., Van IJzendoorn, H.W., Van Lieshout, C.F., & Hartup, W.W. (1992). Heterogeneity among peer rejected boys: Subtypes and stabilities. *Child Development*, **63**, 893-905.

Coie, J.D., & Dodge, K.A. (1983). Continuities and changes in children's social status: A five year longitudinal study. *Merrill-Palmer Quarterly*, **29**, 261-282.

Coie, J.D., Dodge, K.A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, **18**, 557-570.

DeRosier, M.E., Kupersmidt, J.B., & Patterson, C.J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development*, **65**, 1799-1813.

Dunnington, M.J. (1957). Investigation of areas of disagreement in sociometric measurement of preschool children. *Child Development*, **28**, 93- 102.

French, D.C. (1988). Heterogeneity of peer-rejected boys: Aggressive and nonaggressive subtypes. *Child Development*, **59**, 976-985.

French, D.C. (1990). Heterogeneity of peer-rejected girls. *Child Development*, **61**, 2028-2031.

Gronlund, N.E., & Anderson, L. (1957). Personality characteristics of socially accepted, socially neglected and socially rejected junior high school pupils. *Educational Administration and Supervision*, **43**, 329-338.

Maassen, G.H., Akkermans, W., & Van der Linden, J.L. (1996). Two-dimensionsional Sociometric Status Determination with Rating Scales. *Small Group Research*, **27**, 56-78.

Maassen, G.H., & Landsheer, J.A. (1996). *Manual for SSRAT 2.0; A Program for Two-dimensional Sociometric Status Determination with Rating Scales*. Utrecht, the Netherlands: FSW/ISOR Publications, MS 96-2, Utrecht University.

Maassen, G.H., Van der Linden, J.L., & Akkermans, W. (1997). Nominations, Ratings and the Dimensions of Sociometric Status. *International Journal of Behavioral Development*, **21**, 179-199.

Maassen, G.H., Van der Linden, J.L., Goossens, F.A., & Bokhorst, J. (forthcoming). A ratings-based approach to two-dimensional sociometric status determination. In A.H. Cillessen & W.M. Bukowski (eds.), *New Directions in Child Development: Sociometry, part 2*. San Francisco: Jossey Bass.

Newcomb, A.F., & Bukowski, W.M. (1983). Social impact and social preference as determinants of children's peer group status. *Developmental Psychology*, **19**, 856-867.

Newcomb, A.F., Bukowski, W.M., & Pattee, L. (1993). A meta-analytic review of popular, rejected, neglected, controversial and average sociometric status. *Psychological Bulletin*, **113**, 99-128.

Peery, J.C. (1979). Popular, amiable, isolated, rejected: A reconceptualization of sociometric status in preschool children, *Child Development*, **50**, 1231-1234.

Putallaz, M., & Gottman, J.M. (1981). Social skills and group acceptance. In S.R. Asher & J.M. Gottman (Eds.), *The development of children's friendships* (pp.116-149). Cambridge, England: Cambridge University Press.

Ten Brink, P.W.M. (1985). *De gegeneraliseerde binomiale verdeling als alternatief voor de sociometrische status berekening volgens het probabiliteitsmodel* (The generalized binominal distribution as an alternative for sociometric status determination according to the probability model). Nijmegen, The Netherlands: Unpublished master's thesis, Psychologisch Laboratorium, Nijmegen University.

Terry, R., & Coie, J.D. (1991). A Comparison of Methods for Defining Sociometric Status Among Children. *Developmental Psychology*, **27**, 867-880.